Sharpness of the Satisfiability Threshold for Non-Uniform Random $\boldsymbol{k}$-SAT

Tobias Friedrich[1 [0000−0003−0076−6308]] and Ralf Rothenberger[1 [0000−0002−4133−2437]]

Hasso Plattner Institute, Potsdam, Germany
firstname.lastname@hpi.de

Abstract We study non-uniform random k-SAT on n variables with an arbitrary probability distribution $\boldsymbol{p}$ on the variable occurrences. The number $t = t(n)$ of randomly drawn clauses at which random formulas go from *asymptotically almost surely (a. a. s.)* satisfiable to *a. a. s.* unsatisfiable is called the *satisfiability threshold*. Such a threshold is called *sharp* if it approaches a step function as n increases. We show that a threshold $t(n)$ for random k-SAT with an ensemble $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ of arbitrary probability distributions on the variable occurrences is sharp if $\|\boldsymbol{p}_n\|_2^2 = \mathcal{O}_n\left(t^{-\frac{2}{k}}\right)$ and $\|\boldsymbol{p}_n\|_\infty = o_n\left(t^{-\frac{k}{2k-1}} \cdot \log^{-\frac{k-1}{2k-1}} t\right)$.

This result generalizes Friedgut's sharpness result from uniform to non-uniform random k-SAT and implies sharpness for thresholds of a wide range of random k-SAT models with heterogeneous probability distributions, for example such models where the variable probabilities follow a power-law distribution.

1 Introduction

One of the most thoroughly researched topics in theoretical computer science is Satisfiability of Propositional Formulas (SAT). It was one of the first problems shown to be NP-complete by Cook [17] and, independently, by Levin [34]. Furthermore, SAT stands at the core of many results of modern complexity theory, like NP-completeness proofs [33] or lower bounds on runtime assuming the (Strong) Exponential Time Hypothesis [12, 18, 30, 31].

Additional to its importance for theoretical research, Propositional Satisfiability also has practical applications. Many practical problems can be transformed into SAT formulas, for example hard- and software verification, automated planning, and circuit design. Such SAT formulas arising from practical and industrial problems are commonly referred to as *industrial SAT instances*. Surprisingly, even large industrial SAT instances with millions of variables can often be solved efficiently by state-of-the-art SAT solvers. This suggests that these instances have a structure which makes them easier to solve than the theoretical worst-case.

Uniform Random SAT and the satisfiability threshold conjecture: In order to study the average-case complexity of Satisfiability, one can generate a formula Φ at random in conjunctive normal form (CNF) with n variables and m clauses. To this end, we assume to have a probability distribution over all formulas with those properties. If the probability distribution is uniform, we will also refer to the model as *uniform random k-SAT*.

One of the most prominent questions related to uniform random k-SAT is trying to prove the satisfiability threshold conjecture. The *satisfiability threshold conjecture* states that for a formula Φ drawn uniformly at random from the set of all k-CNFs with n variables and m clauses, there is a real number r_k such that

$$\lim_{n \to \infty} \Pr\{\Phi \text{ is satisfiable}\} = \begin{cases} 1 & m/n < r_k; \\ 0 & m/n > r_k. \end{cases}$$

For $k = 2$, Chvatal and Reed [13] and, independently, Goerdt [28] proved that $r_2 = 1$. For $k \geqslant 3$, explicit upper and lower bounds have been derived, e. g., $3.52 \leqslant r_3 \leqslant 4.4898$ [19,29,32]. Additionally, the cavity method from statistical mechanics [35] was used to suggest a numerical estimate of $r_3 \approx 4.26$. Coja-Oghlan and Panagiotou [14,15] derived a bound (up to lower order terms) for $k \geqslant 3$ with $r_k = 2^k \log 2 - \frac{1}{2}(1 + \log 2) \pm o_k(1)$. Recently, Ding, Sly, and Sun [20] proved the exact position of the threshold for sufficiently large values of k.

One goal of showing the conjecture is to rigorously connect or disconnect threshold behavior to the average hardness of solving instances. For uniform random k-SAT for example, the on average hardest instances are concentrated around the threshold [36]. However, the conjecture and a lot of related work only consider formulas that are drawn uniformly at random. But what happens if the formulas are drawn according to a different probability distribution?

Non-Uniform Random SAT: There is a substantial body of work which analyzes the satisfiability threshold in different SAT models, like regular random k-SAT [8, 9, 16, 43], random geometric k-SAT [11] and $2 + p$-SAT [1, 37–39]. However, these models are not motivated by trying to model or understand the properties of industrial instances.

One property of industrial instances is community structure [7], i. e. some variables have a bias towards appearing together in clauses. It is clear by definition that such a bias does not exists in uniform random k-SAT. The Community Attachment Model by Giráldez-Cru and Levy [26] creates random formulas with clear community structure. Yet, the work of Mull et al. [40] shows that instances generated by this model have exponentially long resolution proofs with high probability, making them hard for CDCL on average.

Another important property of industrial instances is their degree distribution. The degree distribution of a formula Φ is a function $f\colon \mathbb{N} \to \mathbb{N}$, where $f(x)$ denotes the number of different Boolean variables that appear x times in Φ (negated or unnegated). In uniform random k-SAT this distribution is binomial, but it has been found out that the degree distribution of many families of industrial instances follows a power-law [5, 10]. This means that $f(x)/n \sim x^{-\beta}$, where β is a constant intrinsic to the instance. To help close the gap between the degree distribution of uniform random and industrial instances, Ansótegui et al. [5] proposed a power-law random SAT model. Empirical studies [3–6] found that SAT solvers that are specialized in industrial instances also perform better on power-law formulas than on uniform random formulas. However, it looks like a power-law degree distribution alone makes instances a bit easier to solve, but not actually "easy": median runtimes around the threshold still look like they scale exponentially for several state-of-the-art solvers [25].

Recently, Giráldez-Cru and Levy [27] also introduced the popularity-similarity model, which incorporates both power-law degree distribution and community structure. Like most other models inspired by industrial instances it lacks theoretical work regarding the satisfiability threshold.

In this work we want to consider a generalization of the power-law random SAT model by Ansótegui et al. [5]. Our model allows instances with *any* given ensemble of variable distributions, instead of just power laws: The variables of each clause are drawn with a probability proportional to the n-th distribution in the ensemble, then they are negated independently with a probability of $1/2$ each. Let $\mathcal{D}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m\right)$ be such a model with a variable distribution ensemble $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$, where m clauses of length k over n variables are drawn. We call this the *clause-drawing* model. If we draw clauses in such a way, the variable probability distribution also defines a probability distribution over k-clauses. Instead of drawing exactly m k-clauses over n variables, one can now imagine flipping a coin for each possible k-clause and taking the clause into the formula with the clause probability multiplied with a certain scaling factor s. By doing so, the expected number of clauses in the formula will be exactly s. We will denote this model by $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s\right)$ and call it the *clause-flipping* model.

Although $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s\right)$ and $\mathcal{D}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m\right)$ cannot represent industrial instances accurately, they might still give us some insights into which properties of real-world instances make them easy to solve. The one property our models provide is degree distribution. They allow us to look at the connection between degree distribution and hardness in an average-case scenario. As one of the steps in analyzing this connection, we would like to find out for which ensembles of variable probability distributions an equivalent of the satisfiability threshold conjecture holds in non-uniform random k-SAT. To see which ingredients we need to prove the conjecture and which of these ingredients this work provides, we first have to introduce the concept of threshold functions formally.

Threshold functions: Formally, due to [23] a threshold for a monotone property P is defined as follows in the classical context of uniform probability distributions: Let $p \in [0,1]$ and let $V = \{0,1\}^N$ be endowed with the product measure $\mu_p(\cdot)$: for $x \in V$ define $\mu_p(x) = p^{\sum x_i}(1-p)^{N-\sum x_i}$, and, for $W \subseteq V$, $\mu_p(W) = \sum_{x\in W} \mu_p(x)$. Now let $P = P(n)$ be the family of properties. $p^* = p^*(n)$ is an *asymptotic threshold function* for $P(n)$ if for every $p = p(n)$

$$\lim_{n\to\infty} \mu_p(P) = \begin{cases} 0, & \text{if } p \ll p^* \\ 1, & \text{if } p \gg p^*. \end{cases}$$

Here $\ll$ and $\gg$ denote "asymptotically smaller" and "asymptotically bigger" respectively.

Intuitively, a *sharp threshold* means that the change in probability around the threshold becomes steeper and steeper as the problem size increases, converging to a step function as n tends to infinity. Formally, we say that $P(n)$ has a *sharp threshold* if there exists a function $p^* = p^*(n)$ such that for every constant $\varepsilon > 0$ and for every $p = p(n)$

$$\lim_{n\to\infty} \mu_p(P) = \begin{cases} 0, & \text{if } p \leqslant (1-\varepsilon)p^* \\ 1, & \text{if } p \geqslant (1+\varepsilon)p^*. \end{cases}$$

Otherwise we call a threshold *coarse*. The region of p where the limit of $\mu_p(P)$ is bounded away from zero and one is called the *threshold interval.*

Note, that this definition only holds for satisfiability in the uniform clause-flipping model. In the case of the uniform clause-drawing model, the sharpness of the threshold is defined the same way, but with respect to m (or $r = m/n$) instead of p on the appropriate probability space.

Proving the satisfiability threshold conjecture: In terms of threshold functions, the conjecture states that there is a sharp threshold for satisfiability at $m = r_k \cdot n$ and the constant r_k is the same for a fixed k and all sufficiently large n. For $k = 2$, Chvatal and Reed [13] and Goerdt [28] proved the conjecture and showed that $r_2 = 1$. However, random 2-SAT is easier to analyze than random k-SAT and their techniques do not work for bigger values of k. For uniform random k-SAT the "recipe" for proving the conjecture is as follows:

1. Show the existence of an asymptotic threshold function, i. e. show constant lower and upper bounds on r_k.
2. Prove that the threshold is sharp. In 1999 Friedgut [22] showed that the satisfiability threshold for uniform random k-SAT is sharp, although its location is not known exactly for all values of k. However, his result does not prove that r_k is the same for a fixed k and all sufficiently large values of n. Friedgut's proof relies on knowing the asymptotic threshold function.
3. Derive the actual constant r_k and that the threshold is sharp around it. Ding et al. [20] were the first to prove the exact value of r_k for values of k bigger than 2. Their proof relies on the result of Friedgut.

The goal of this paper will be to show the second ingredient for proving the satisfiability threshold conjecture for non-uniform random k-SAT, sharpness of the satisfiability threshold. In addition to being a stepping stone to proving the conjecture, sharpness of the threshold is of some independent interest, since a coarse threshold implies that there is a local property which approximates satisfiability or unsatisfiability. For random SAT this means that with constant probability instances have a constant-sized unsatisfiable subformula, making a lot of instances very easy to solve even around the threshold. Moreover, some of the techniques we use could also be used to analyze more sophisticated models, e.g. the popularity-similarity model [27], which was used in the 2017 SAT Competition.

Our results: We study the sharpness of the satisfiability threshold for non-uniform random k-SAT and identify sufficient conditions on the variable probability distribution which imply a sharp threshold. Therefore, this work provides the second ingredient for proving a version of the satisfiability threshold conjecture for the non-uniform models $\mathcal{D}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m\right)$ and $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s\right)$ introduced earlier. In the context of these models, the classical result of Friedgut [22] reads as follows:

Theorem 1.1 (by [22]). *For all $n \in \mathbb{N}$ let $\boldsymbol{p}_n = (1/n, 1/n, \ldots, 1/n)$ be a variable probability distribution on n variables. If there is an asymptotic satisfiability threshold $m_c = t(n)$ on $\mathcal{D}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m\right)$, then satisfiability has a* sharp *threshold on $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s\right)$ with respect to s, and a* sharp *threshold on $\mathcal{D}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m\right)$ with respect to m.*

Our main theorem extends this to non-uniform random k-SAT:

Theorem 3.2. *Let $k \geqslant 2$, let $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ be an ensemble of variable probability distributions on n variables each and let $s_c = t(n)$ be an asymptotic satisfiability threshold for $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s\right)$ with respect to s. If $\|\boldsymbol{p}_n\|_\infty = o\left(t^{-\frac{k}{2k-1}} \cdot \log^{-\frac{k-1}{2k-1}} t\right)$ and $\|\boldsymbol{p}_n\|_2^2 = \mathcal{O}\left(t^{-2/k}\right)$, then satisfiability has a sharp threshold on $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s\right)$ with respect to s.*

Furthermore, we show that the same also holds for the clause-drawing model of non-uniform random k-SAT if the asymptotic threshold is not constant.

Theorem 3.3. *Let $k \geqslant 2$, let $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ be an ensemble of variable probability distributions on n variables each and let $m_c = t(n) = \omega(1)$ be the asymptotic satisfiability threshold for $\mathcal{D}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m\right)$ with respect to m. If $\|\boldsymbol{p}_n\|_\infty = o\left(t^{-\frac{k}{2k-1}} \cdot \log^{-\frac{k-1}{2k-1}} t\right)$ and $\|\boldsymbol{p}_n\|_2^2 = \mathcal{O}\left(t^{-2/k}\right)$, then satisfiability has a sharp threshold on $\mathcal{D}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m\right)$ with respect to m.*

Our results actually state that the threshold is sharp for a certain, fixed value of n in the sense that the probability function for unsatisfiability is $o(1)$ if $s = (1-\varepsilon) \cdot s_c$ (or $m = (1-\varepsilon) \cdot m_c$) and $1 - o(1)$ if $s = (1+\varepsilon) \cdot s_c$ (or $m = (1+\varepsilon) \cdot m_c$). It is still possible that the function behaves differently for higher n due to the changing number of variables and probabilities. Nevertheless, Friedgut's original result also only asserts sharpness for a certain, fixed value of n. This is also the reason why the sharp threshold result does not automatically prove the satisfiability threshold conjecture: There could be different sharp threshold functions (including leading constant factors) for different values of n. For example, there could be some strange oscillations of the function.

Techniques: The proof of the main theorem uses Bourgain's Sharp Threshold Theorem in the version from O'Donnell's book [42]. In general, it follows the lines of Friedgut's proof of sharpness for the threshold of uniform random k-SAT [22].

However, we have to generalize Friedgut's results, like showing that no short unsatisfiable subformula can exist with sufficiently high probability. Furthermore, his lemma to bound the maximum slope of the probability for a monotone property at the threshold cannot be applied anymore, even in a more general form. Instead, we use the maximum slope that is implied by assuming a coarse threshold. Also, we had to adapt Friedgut's coverability lemma when considering non-uniform random k-SAT. In his work, a quasi-unsatisfiable subformula can spawn a constant number of clauses of length $k-1$. Now a quasi-unsatisfiable subformula can spawn clauses of any length $l \leqslant k$. Furthermore, there can now be more than a constant number of spawned clauses.

Please note that due to space limitations, we only provide proof sketches for our results. The full proofs can be found in the full version of the paper.

2 Preliminaries

We analyze random k-SAT on n variables and m clauses. We denote by $X_1, \ldots, X_n$ the Boolean variables. A clause is a disjunction of k literals $\ell_1 \vee \ldots \vee \ell_k$, where each literal assumes a (possibly negated) variable. For a literal ℓ_i let $|\ell_i|$ denote the variable of the literal. A formula $\varPhi$ in conjunctive normal form is a conjunction of clauses $c_1 \wedge \ldots \wedge c_m$. We conveniently interpret a clause c both as a Boolean formula and as a set of literals. We say that $\varPhi$ is satisfiable if there exists an assignment of variables $X_1, \ldots, X_n$ such that the formula evaluates to 1. Now let $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ be an ensemble of probability distributions, where $\boldsymbol{p}_n = (p_{n,1}, p_{n,2}, \ldots, p_{n,n})$ is a probability distribution over n variables with $\Pr(X = X_i) = p_{n,i} =: p_n(X_i)$.

Definition 2.1 (Clause-Drawing Random k-SAT). *Let m, n, k be given, and consider any ensemble of probability distributions $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$, where $\boldsymbol{p}_n = (p_{n,1}, p_{n,2}, \ldots, p_{n,n})$ is a probability distribution over n variables with $\sum_{i=1}^{n} p_{n,i} = 1$. The* clause-drawing random k-SAT *model $\mathcal{D}(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m)$ constructs a random SAT formula $\varPhi$ by sampling m clauses independently at random. Each clause is sampled as follows:*

1. *Select k variables independently at random from the distribution $\boldsymbol{p}_n$. Repeat until no variables coincide.*
2. *Negate each of the k variables independently at random with probability $1/2$.*

The clause-drawing random k-SAT model is equivalent to drawing each clause independently at random from the set of all k-clauses which contain no variable more than once. The probability to draw a clause c over n variables is then

$$q_c := \frac{\prod_{\ell\in c} p_n(|\ell|)}{2^k \sum_{J\in\mathcal{P}_k(\{1,2,\ldots,n\})} \prod_{j\in J} p_{n,j}}, \tag{2.1}$$

where $\mathcal{P}_k(\cdot)$ denotes the set of cardinality-k elements of the power set. The factor 2^k in the denominator comes from the different possibilities to negate variables. Note that $k! \sum_{J\in\mathcal{P}_k(\{1,2,\ldots,n\})} \prod_{j\in J} p_{n,j}$ is the probability of choosing a k-clause that contains no variable more than once. To see that this probability is almost 1 for most distributions, we apply the generalized birthday paradox from [2]. Thereby, the probability that a k-clause sampled on n variables has collisions is at most $\frac{1}{2}k^2\|\boldsymbol{p}_n\|_2^2$; so for $\|\boldsymbol{p}_n\|_2^2 = o(1)$ and constant k we obtain that the probability to draw a specific clause over n variables consisting of variables $X \in S$ is

$$q_c = C\frac{k!}{2^k}\prod_{X\in S} p_n(X), \tag{2.2}$$

where we define $C := 1/\left(\sum_{J\in\mathcal{P}_k(\{1,2,\ldots,n\})}\prod_{j\in J} p_{n,j}\right) = \left(1+\Theta\left(\|\boldsymbol{p}_n\|_2^2\right)\right)$. This effectively hides the denominator of equation (2.1) in C and makes clause probabilities easier to handle. We will later see that this is always the case in the variable probability distributions we consider.

We can now define the coin-flipping equivalent of non-uniform random k-SAT, which we will label *clause-flipping random k-SAT*.

Definition 2.2 (Clause-Flipping Random k-SAT). *Let s, n, k be given, and consider any ensemble of probability distributions $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$, where $\boldsymbol{p}_n$ is a probability distribution over n variables with $\sum_{i=1}^n p_i = 1$. The* clause-flipping random k-SAT *model $\mathcal{F}(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s)$ constructs a random SAT formula Φ over n variables by independently flipping a coin for each of the $\binom{n}{k}2^k$ possible k-clauses. The coin flip for a clause c is a success with probability*

$$q_{n,c}(s) := \min\left(s\cdot q_{n,c}, 1\right) = \min\left(s\cdot\frac{\prod_{\ell\in c} p_n(|\ell|)}{2^k\sum_{J\in\mathcal{P}_k(\{1,2,\ldots,n\})}\prod_{j\in J} p_{n,j}}, 1\right).$$

If successful, the clause is added to the random formula.

Lemma 2.1 relates the two models to each other and will be used throughout the paper. Note that in the lemma the clause probabilities do not necessarily have to be products of variable probabilities! Its proof is a simple exercise.

Lemma 2.1. *Given a clause-flipping model $\mathcal{F}$ with clause probabilities $\boldsymbol{q} = (q_i)_{i\in[n]}$ and a clause-drawing model $\mathcal{D}$ with clause probabilities $\boldsymbol{q'} = (q'_i)_{i\in[n]}$ so that $q'_i = \frac{q_i/(1-q_i)}{\sum_{j\in[n]} q_i/(1-q_i)}$, then for all $l \in \mathbb{N}$ and all events $\mathcal{E}$ it holds that*

$$\Pr_{\mathcal{F}}\left(\,\mathcal{E} \mid \{l \text{ clauses flipped}\}\,\right) = \Pr_{\mathcal{D}}\left(\,\mathcal{E} \mid \{l \text{ different clauses drawn}\}\,\right).$$

3 Sharpness of the Threshold

In Section 3.1 we establish a notion of asymptotic and sharp thresholds in the context of non-uniform probability distributions. In Section 3.2 we relate this notion of sharpness to Bourgain's Sharp Threshold Theorem. In Section 3.3 we prove the sharpness of the threshold in $\mathcal{F}(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s)$ with the help of the Sharp Threshold Theorem. Finally, in Section 3.4 we relate $\mathcal{D}(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m)$ to $\mathcal{F}(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s)$ in such a way that the sharpness of the satisfiability threshold carries over.

3.1 Non-Uniform Sharpness

We want to generalize the definitions for uniform probability distributions to non-uniform probability distributions.

For the clause-drawing random k-SAT model, we can use the same concepts of asymptotic and sharp thresholds with respect to m as in the uniform case.

For the clause-flipping random k-SAT model, the first thing we notice is that we cannot define the thresholds with respect to p anymore since the clause probabilities are now non-uniform. Instead, we want to define the thresholds with respect to s, the scaling factor of the probability space. This will allow us to relate the two models in subsection 3.4.

Unless stated otherwise, we will concentrate on $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s\right)$ to establish the result in this model first. We now encode formulas as vectors $x \in \{0,1\}^N$, where $N := \binom{n}{k}2^k$ is the number of different k-clauses over n variables. If a clause is chosen to be in the formula, we set its variable to -1, otherwise we set it to 1. With this encoding of k-CNFs in mind, we can define a function $f\colon \{-1,1\}^N \to \{-1,1\}$, which returns -1 if the encoded k-CNF is unsatisfiable and 1 otherwise. It is easy to see that f is monotone in the sense that $f(x) \leqslant f(y)$ whenever $x \leqslant y$ coordinate-wise. This is the case, since setting a coordinate from -1 to 1 is equivalent to removing a clause from the encoded formula. By doing so, a satisfiable formula cannot be made unsatisfiable, i. e. the value of f can only change from -1 to 1, but not the other way around. This encoding is from O'Donnell's book [42] and makes the application of Bourgain's Sharp Threshold Theorem later in the paper easier.

We can now formally describe the product probability space of $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s\right)$ with the notation of O'Donnell. Given a variable probability distribution $\boldsymbol{p}_n = (p_{n,i})_{i=1,\ldots,n}$, the derived clause probability distribution $\boldsymbol{q}_n = (q_{n,i})_{i=1,\ldots,N}$, and the scaling factor s, we define our product space to be $(\Omega, \pi) := \left(\{-1,1\}^N, \pi_1 \times \pi_2 \times \ldots \times \pi_N\right)$ with $\pi_i(-1) = q_{n,i}(s)$ and $\pi_i(1) = 1 - q_{n,i}(s)$ for $i = 1, 2, \ldots, N$. We let $\mu_{\boldsymbol{p}_n,s}$ denote the product probability measure, i.e. for $x \in \Omega$

$$\mu_{\boldsymbol{p}_n,s}(x) = \prod_{i=1}^{N} \pi_i(x_i) = \prod_{i\in[N]\colon x_i=-1} q_{n,i}(s) \prod_{i\in[N]\colon x_i=1} (1 - q_{n,i}(s)).$$

For $S \subseteq \Omega$ we define $\mu_{\boldsymbol{p}_n,s}(S) = \sum_{x\in S} \mu_{\boldsymbol{p}_n,s}(x)$. We will use the shorthand notation μ instead of $\mu_{\boldsymbol{p}_n,s}$ if the probability measure is clear from context. Furthermore, for an N-element vector $x = (x_1, x_2, \ldots, x_N)$ and a subset $T \subseteq [N]$ let $x_T = (x_i)_{i\in T}$ denote the *restriction of x to T*.

The following statement shows the relation between coarseness of a property's threshold and the derivative of its probability function. The uniform equivalent of the statement holds due to Friedgut [22], but we can show that it also holds in the non-uniform case. The proof of the statement is a simple application of the mean value theorem.

Lemma 3.1. *If a threshold is coarse, then there is a point s^* in the threshold interval, where $s^* \cdot \frac{d\mu_{\boldsymbol{p}_n,s}(f)}{ds}|_{s=s^*} \leqslant K$ for some constant K.*

3.2 Influence and Bourgain's Sharp Threshold Theorem

Bourgain's Sharp Threshold Theorem will make use of the total influence of a Boolean function f. Intuitively, the *influence* $\mathbf{Inf}_i[f]$ of a function f describes the probability that the value of the i-th coordinate influences the function value. The *total influence* $\mathbf{I}[\,f\,]$ of a function f is the sum of the influence values for all coordinates. Both, $\mathbf{Inf}_i[f]$ and $\mathbf{I}[\,f\,]$ depend on the probability distribution π, but we will omit this dependence if it is clear from context. The following definition from [42] formalizes our intuitive one.

Definition 3.1. *[Influence Function] Let $f \in L^2(\Omega, \pi)$ be $\{-1,1\}$-valued with $\Omega = \{-1,1\}^N$ and $\pi = \pi_1 \times \ldots \times \pi_N$. The* influence *of the i-th coordinate is $\mathbf{Inf}_i[f] = \mathop{\mathbb{E}}_{x\sim\pi}[\,f(x)(L_i f)(x)\,]$*[1]*, where $L_i f = f - E_i f$ and $E_i f(y) = \mathop{\mathbb{E}}_{\boldsymbol{y_i}\sim\pi_i}[\,f(y_1, y_2, \ldots, y_{i-1}, \boldsymbol{y_i}, y_{i+1} \ldots, y_{N-1}, y_N)\,]$. The* total influence *of f is $\mathbf{I}[\,f\,] = \sum_{i=1}^{n} \mathbf{Inf}_i[f]$.*

The following corollary relates this notion of influence to the notion of coarseness due to Friedgut, more precisely to $\frac{d\mu_{\boldsymbol{p}_n,s}(P)}{ds}s = \frac{d\mu_{\boldsymbol{p}_n,s}(\{x\in\Omega|f(x)=-1\})}{ds}s$. Its proof is a relatively simple exercise.

Corollary 3.1. *Let $f \in L^2\left(\Omega = \{-1,1\}^N, \pi = \pi_1 \times \pi_2 \times \ldots \times \pi_N\right)$ be $\{-1,1\}$-valued, monotone, and non-constant. For $s < \left(\max_{i\in[N]}(q_{n,i})\right)^{-1}$ it holds that*

$$\mathbf{I}[\,f\,] \leqslant 4 \cdot \frac{d\mu_{\boldsymbol{p}_n,s}(\{x \in \Omega \mid f(x) = -1\})}{ds} s. \tag{3.1}$$

To prove our main theorem, we will use the Sharp Threshold Theorem by Bourgain [22] in O'Donnell's version [42]. The theorem states that, if a monotone property P has a coarse threshold, and therefore small influence, then there are local structures which approximate this property. The following is a formal definition of these structures.

Definition 3.2. *[τ-booster] Let $f\colon \Omega \to \{-1,1\}$. For $T \subseteq [N]$, $y \in \Omega$, and $\tau > 0$, we say that the restriction y_T is a τ-*booster *if $\mathop{\mathbb{E}}_{x\sim\pi}[\,f \mid x_T = y_T\,] \geqslant \mathbb{E}[\,f\,] + \tau$. If $\tau < 0$, we say that y_T is a τ-booster if $\mathop{\mathbb{E}}_{x\sim\pi}[\,f \mid x_T = y_T\,] \leqslant \mathbb{E}[\,f\,] - |\tau|$.*

The Sharp Threshold Theorem is stated as follows:

Theorem 3.1. *[Bourgain's Sharp Threshold Theorem] Let $f \in L^2(\Omega, \pi)$ be $\{-1,1\}$-valued and non-constant with $\mathbf{I}[\,f\,] \leqslant K$ for a constant K. Then there is some τ (either negative or positive) with $|\tau| \geqslant \mathbf{Var}[\,f\,] \cdot \exp(-\mathcal{O}(\mathbf{I}[\,f\,]^2/\mathbf{Var}[\,f\,]^2))$ such that*

$$\Pr_{x\sim\pi}\left(\exists\; T \subseteq [n],\; |T| \leqslant \mathcal{O}\left(\frac{\mathbf{I}[\,f\,]}{\mathbf{Var}[\,f\,]}\right) \text{ such that } x_T \text{ is a } \tau\text{-booster}\right) \geqslant |\tau|.$$

[1] In the paper we let $x \sim \pi$ denote that the random variable x is drawn from the probability distribution with density π.

This Theorem seems to be specific to probability spaces with uniform probability distributions. However, O'Donnell states that Theorem 3.1 in the version with arbitrary product probability spaces also holds. We verify this claim in the full version of the paper. Furthermore, by carefully checking the proof of the theorem, one can see that the asymptotic values and the bases for the exponential terms can actually be substituted by appropriately chosen exact expressions. Also, it has to be noted that Müller [41] already showed that a version of Bourgain's original theorem also holds for arbitrary product probability spaces.

3.3 Proof of Sharpness for Non-Uniform Random k-SAT

This subsection will be dedicated to proving our main theorem:

Theorem 3.2. *Let $k \geqslant 2$, let $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ be an ensemble of variable probability distributions on n variables each and let $s_c = t(n)$ be an asymptotic satisfiability threshold for $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s\right)$ with respect to s. If $\|\boldsymbol{p}_n\|_\infty = o\left(t^{-\frac{k}{2k-1}} \cdot \log^{-\frac{k-1}{2k-1}} t\right)$ and $\|\boldsymbol{p}_n\|_2^2 = \mathcal{O}\left(t^{-2/k}\right)$, then satisfiability has a sharp threshold on $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s\right)$ with respect to s.*

The proof closely follows the one by Friedgut for uniform random k-SAT [22]. We assume toward a contradiction that the threshold is coarse. Then the Sharp Threshold Theorem tells us that there have to be so-called "boosters" of constant size that appear with constant probability in the random formula. These boosters have the property that conditioning on their existence *boosts* the probability of the random formula to be unsatisfiable by at least an additive constant.

One kind of booster are unsatisfiable subformulas of constant size. Conditioning on these would boost the probability to be unsatisfiable to one. We rule these out by showing that they do not appear with constant probability.

Then, we consider subformulas, which give the second highest boost: maximally quasi-unsatisfiable subformulas. These are subformulas which have only *one* satisfying assignment for the variables appearing in them and adding any new clause over those variables makes them unsatisfiable. We want to show that these cannot boost the probability of a formula to be unsatisfiable by a constant.

Again toward a contradiction, we assume, that conditioning on a maximally quasi-unsatisfiable subformula T is enough to boost the unsatisfiability probability by a constant. First, we prove that conditioning on T is equivalent to adding a number of clauses of size shorter than k to the random formula over variables not appearing in T. Then, we use a version of Friedgut's coverability lemma to show that, if adding these clauses of size smaller than k makes the random formula unsatisfiable with constant probability, then so does adding $o(t)$ clauses of size k. We prove that this probability is dominated by the probability to make the original random formula unsatisfiable for a slightly bigger scaling factor. However, due to the assumption of a coarse threshold, the slope of the probability function for unsatisfiability has to be small at one point in the threshold interval. If we consider this point, the probability to make the original random formula unsatisfiable cannot be increased by a constant with our slightly increased scaling factor. This contradicts our assumption that the probability is boosted by a constant in the first place. Therefore, quasi-unsatisfiable subformulas cannot be boosters.

After showing this, every less restrictive subformula cannot be a booster either. That means, the only possible boosters are unsatisfiable subformulas, which we ruled out already. Therefore, the implication of the Sharp Threshold Theorem does not hold, which contradicts the assumption of a coarse threshold.

Now we are ready to prove our main theorem.

Application of the Sharp Threshold Theorem We know that the asymptotic threshold is at a scaling factor $s = \Theta(t(n))$. A threshold due to our definition always has to be $t = \Omega(1)$. Otherwise the expected number of clauses would be $O(t) = o(1)$, leading to a probability of $1 - o(1)$ of having an empty, and thereby satisfiable, formula due to Markov's inequality. We can thus assume that $C = \left(1 + o\left(t^{-1/k}\right)\right)$ due to equation (2.2).

To prove Theorem 3.2 we assume that the threshold is coarse. Due to Lemma 3.1 this implies that $\frac{d\mu_{\boldsymbol{p}_n,s}(f)}{ds}s \leqslant K$ for some constant K and some s in the threshold interval. Let us call this scaling factor s_c. Note that $s_c = \Theta(t)$, since s_c is in the threshold interval and t is an asymptotic threshold function. Due to Corollary 3.1 this means $\mathbf{I}\left[\,f\,\right] \leqslant 4 \cdot K$. For the corollary to hold, we have to assure $s_c < \left(\max_{i\in[N]}(q_{n,i})\right)^{-1}$. This follows due to our assumption

$$p_{n,\max} := \|\boldsymbol{p}_n\|_\infty = o\left(t^{-\frac{k}{2k-1}} \cdot \log^{-\frac{k-1}{2k-1}} t\right) = o\left(t^{-1/k}\right),$$

which implies

$$q_{n,\max}(s_c) := \max_{i\in[N]}\left(q_{n,i}(s_c)\right) = s_c \cdot \mathcal{O}\left(p_{n,\max}^k\right) = o(1). \tag{3.2}$$

Since f is $\{-1, 1\}$-valued it holds that $\mathbb{E}\left[\,f\,\right] = 1 - 2 \cdot \mu_{\boldsymbol{p}_n,s_c}(f)$ and $\mathbf{Var}\left[\,f\,\right] = 4 \cdot \mu_{\boldsymbol{p}_n,s_c}(f)\left(1 - \mu_{\boldsymbol{p}_n,s_c}(f)\right)$. Since we are in the threshold interval, it holds that $\mu_{\boldsymbol{p}_n,s_c}(f)$ is constant and so are $\mathbb{E}\left[\,f\,\right]$ and $\mathbf{Var}\left[\,f\,\right]$.

Now we can use Theorem 3.1 to see that, at least with constant probability τ, our formulas have a subformula (or lack thereof) consisting of at most $\mathcal{O}(K) = \mathcal{O}(1)$ clauses, so that conditioning on the existence (or non-existence) of these clauses increases (or decreases) the probability that our random k-CNFs are unsatisfiable by at least $\tau/2$. The subformulas with these properties are the boosters. The theorem actually allows us to choose appropriate specific constants for τ and the upper bound on $|T|$.

Since the property of being unsatisfiable is monotone, it is not beneficial to forbid some clauses and demand others. We can therefore concentrate on the cases of either only forbidding or only enforcing clauses in our boosters. The following lemma shows that it suffices to concentrate on enforcing boosters. The idea is that every constant-sized subset of clauses a. a. s. does not exist in the formula, since clause probabilities are $o(1)$. Therefore, conditioning on the non-existence of such a subformula does not change the overall probability by too much.

Lemma 3.2. *Every booster which assumes the non-existence of clauses only boosts the probability to be satisfiable or unsatisfiable by* $o(1)$.

We can now concentrate on conditioning on the *existence* of clauses. Our goal is to show that no constant-sized boosters exist with constant probability.

Unsatisfiable subformulas are too improbable A sure way to boost the probability of being unsatisfiable to one is to condition on the existence of an unsatisfiable subformula. To rule this case out, the next lemma shows that the probability that our formulas have an unsatisfiable subformula of constant size is smaller than any constant τ for sufficiently large n. The proof essentially shows that any minimally unsatisfiable subformula of constant size cannot exist with constant probability. This can be seen from the fact that such subformulas contain each variable in them at least twice and the probability for this can be bounded using $\|\boldsymbol{p}_n\|_2^2$ and $\|\boldsymbol{p}_n\|_\infty$.

Lemma 3.3. *Let* $a, k \in \mathbb{N}$ *be constants and let* $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ *be an ensemble of variable probability distributions. If* $\|\boldsymbol{p}_n\|_\infty = o\left(s^{-1/k}\right)$ *and* $\|\boldsymbol{p}_n\|_2^2 = \mathcal{O}\left(s^{-2/k}\right)$, *then a random formula from* $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s\right)$ *has an unsatisfiable subformula of length at most* a *with probability* $o(1)$.

Maximally quasi-unsatisfiable subformulas provide the second-highest boost Since we ruled out unsatisfiable subformulas as the boosters we are looking for, we now turn our attention to satisfiable subformulas. Let $\varPhi_T$ be the formula encoded by $x_T = (-1)^{|T|}$ and let $V(T) \subseteq \{X_1, \ldots, X_n\}$ be the variables in $\varPhi_T$. Note that $|V(T)|$ is constant since $|T|$ is constant and each clause contains k many variables. We call $\varPhi_T$ *maximally quasi-unsatisfiable (mqu)* if it is satisfiable by only one of the $2^{|V(T)|}$ assignments over its variable set (quasi-unsatisfiable) and if adding any new clause with variables only from $V(T)$ makes it unsatisfiable (maximally satisfiable). The following lemma formalizes a statement by Friedgut [22] that the biggest possible boost any satisfiable subformula can give is achieved by mqu subformulas. The proof of the statement uses the fact that every satisfiable subformula can be extended to a mqu subformula over the same variables. It also uses positive correlation of increasing events [21] and the fact that we have a product probability space.

Lemma 3.4. *For every* $T \subseteq [N]$ *so that* $\varPhi_T$ *is satisfiable, there is a* $T' \supseteq T$ *so that* $\varPhi_{T'}$ *is maximally quasi-unsatisfiable and*

$$\Pr_{x\sim\pi}\left(f(x) = -1 \mid x_{T'} = (-1)^{|T'|}\right) \geqslant \Pr_{x\sim\pi}\left(f(x) = -1 \mid x_T = (-1)^{|T|}\right).$$

The part of the formula containing only variables from the booster is still satisfiable We now turn to analyzing the boost maximally quasi-unsatisfiable subformulas can give. In the end will will show that they cannot boost the unsatisfiability probability by a constant. Lemma 3.4 implies that the same holds for all satisfiable subformulas, thus giving us the desired contradiction.

Let $T \subseteq [N]$ with Φ_T mqu. In order to see how big the boost by such a T can be, we split x into two parts, the part x_S, so that each clause in Φ_S only contains variables from $V(T)$, and the part $x_{\overline{S}}$, in which each encoded clause contains at least one variable from $\overline{V(T)} = \{X_1, \ldots, X_n\} \setminus V(T)$. Let $f(x_S)$ be -1 if Φ_S is unsatisfiable and 1 otherwise. The following lemma asserts that Φ_S can only be unsatisfiable with sub-constant probability. This is the case, because it is very unlikely to flip one of the constant number of clauses that can make the maximally satisfiable booster unsatisfiable.

Lemma 3.5. *It holds that* $\Pr_{x \sim \pi}\left(f(x_S) = -1 \mid x_T = (-1)^{|T|}\right) = o(1)$.

The booster adds shorter clauses to the other part of the formula We can now concentrate on the case that Φ_S is satisfiable. Since Φ_T is maximally unsatisfiable, it holds that $\Phi_S = \Phi_T$, and since Φ_T is quasi-unsatisfiable, Φ_S also only has one satisfying assignment.

We now want to create $x_{\overline{S}}$ under these conditions. To this end, we assume that the variables $V(t)$ take the one assignment that makes Φ_S satisfiable. For a clause containing both variables from $V(T)$ and variables from $\overline{V(T)}$ this means the clause is either satisfied or the variables from $V(T)$ can be eliminated as their literals are all set to false. Effectively, this means that we can have clauses over $\overline{V(T)}$ of length $0 < l < k$. The following lemma makes this statement more precise. Its proof is a simple application of the Markov bound.

Lemma 3.6. *If* $p_{n,\max} = o\left(t^{-\frac{k}{2k-1}} \cdot \log^{-\frac{k-1}{2k-1}} t\right)$, *then a mqu subformula of constant length spawns at most* $D_l = o\left(\left(\frac{t}{\log t}\right)^{\frac{l}{k+l}}\right)$ *clauses of length* $l = 1, \ldots, k-1$ *with probability* $1 - o(1)$.

We now want to create the resulting formula over variables from $\overline{V(T)}$ in two parts. First we create k-clauses over $\overline{V(T)}$ with the usual clause-flipping model, where the clause-probabilities are the same as in $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n \in \mathbb{N}}, s_c\right)$. Then, for each $l \in [k-1]$ we add D_l l-clauses over $\overline{V(T)}$ with the clause-drawing model. The probability q_c to add a clause $c = (\ell_1 \vee \ell_2 \vee \ldots \vee \ell_l)$ of size l is equal to the probability of flipping any clause which contains c and $k - l$ literals negated by the assignment of $\overline{V(T)}$:

$$q_c = C \frac{k! \cdot s_c}{2^k} \prod_{i=1}^{l} p_n(|\ell_i|) \cdot \sum_{J \in \mathcal{P}_{k-l}(V(T))} \prod_{X \in J} p_n(X). \tag{3.3}$$

We can now choose $q'_c = \frac{q_c/(1-q_c)}{\sum_{j \in [n]} q_c/(1-q_c)}$ as the probability to draw clause c. This helps us apply Lemma 2.1 to relate the resulting random formula $\hat{\Phi}$ to our original probability space. Furthermore, the following lemma also uses Lemma 3.6 and the fact that no clauses are drawn twice with probability $1 - o(1)$.

Lemma 3.7. *It holds that*

$$\Pr_{x \sim \pi}\left(f(x) = -1 \wedge f(x_S) = 1 \mid x_T = (-1)^{|T|}\right) \leqslant \Pr\left(\hat{\Phi} \textit{ unsat}\right) + o(1).$$

Shorter clauses can be substituted with $\boldsymbol{k}$-clauses We now want to bound $\Pr(\hat{\Phi}$ unsat$)$. To this end, let $\widetilde{\Phi}$ be the part of $\hat{\Phi}$ only consisting of k-clauses. Let us *assume* $\Pr(\hat{\Phi}$ unsat$) \geqslant \mu_{\boldsymbol{p}_n,s_c}(f) + \delta$ for some constant $\delta > 0$. We know that $\widetilde{\Phi}$ is unsatisfiable with probability at most $\mu_{\boldsymbol{p}_n,s_c}(f)$, since it is drawn from $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s_c\right)$ with the difference that only clauses over $\overline{V(T)}$ are flipped. This implies $\Pr(\hat{\Phi}$ unsat $\wedge\, \widetilde{\Phi}$ sat$) \geqslant \delta$. We now define a more general concept of coverability, analogously to Friedgut [22]. This will allow us to substitute l-clauses with k-clauses while maintaining the probability to make $\widetilde{\Phi}$ unsatisfiable.

Definition 3.3. *Let $D_1, \ldots, D_a \in \mathbb{N}$ and $l_1, \ldots, l_a \in \mathbb{N}$ and let $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_a$ be probability distributions. For $A \subseteq \{0,1\}^n$, we define A to be* $((d_1, l_1, \boldsymbol{q}_1), (d_2, l_2, \boldsymbol{q}_2), \ldots, (d_a, l_a, \boldsymbol{q}_a), \varepsilon)$-coverable*, if the union of d_i subcubes of co-dimension l_i chosen according to probability distribution $\boldsymbol{q}_i$ for $1 \leqslant i \leqslant a$ has a probability of at least ε to cover A.*

In contrast to Friedgut's definition, we allow subcubes of arbitrary co-dimension and with arbitrary probability distributions instead of only subcubes of co-dimension 1 with a uniform distribution. In the context of satisfiability we say that a specific formula (*not* a random formula) Φ is $((d_1, l_1, \boldsymbol{q}_1), \ldots, (d_a, l_a, \boldsymbol{q}_a), \varepsilon)$-coverable if the probability to make it unsatisfiable by adding d_i random clauses of size l_i chosen according to distribution $\boldsymbol{q}_i$ for $i = \underline{1, 2, \ldots} a$ is at least ε in total.

Now let $\boldsymbol{q}_l = (q'_c)_c$ for all clauses c of size l over $\overline{V(T)}$, where q'_c is the clause drawing probability we defined for $\hat{\Phi}$. It holds that with a sufficiently large constant probability $\widetilde{\Phi}$ is $((D_1, 1, \boldsymbol{q}_1), \ldots, (D_{k-1}, k-1, \boldsymbol{q}_{k-1}), \delta)$-coverable. The following lemma shows that formulas with this property are also $((g(n), k, \boldsymbol{q}_k), \delta')$-coverable for some function $g(n) = o\left(t\right)$ and any constant $\delta' < \delta$. Its proof is a more precise and general version of Friedgut's original proof.

Lemma 3.8. *Let $\boldsymbol{q}_k$ be our original clause probability distribution and let all other probability distributions be as described in equation* (3.3) *and let $D_1 \ldots D_{k-1}$ be as defined. If a concrete formula Φ is $((D_1, 1, \boldsymbol{q}_1), \ldots, (D_{k-1}, k-1, \boldsymbol{q}_{k-1}), \delta)$-coverable for some constant $\delta > 0$, it is also $((g(t), k, \boldsymbol{q}_k), \delta')$-coverable for some function $g(t) = o(t)$ for any constant $0 < \delta' < \delta$.*

By substituting shorter clauses with k-clauses we lose at most an arbitrarily small additive constant from the probability $\mu_{\boldsymbol{p}_n,s_c}(f) + \delta$ that $\hat{\Phi}$ is unsatisfiable. Thus, we still have a constant probability bigger than $\mu_{\boldsymbol{p}_n,s_c}(f)$.

Bounding the boost by bounding the slope of the probability function We can now show that instead of adding $g(t)$ k-clauses, we can increase the scaling factor s_c of our original clause-flipping model to achieve the same probability. The proof of the following lemma uses Lemma 2.1 together with a Chernoff-Bound on the number of clauses added in the clause-flipping model.

Lemma 3.9. *For $g'(t) = g(t) + c \cdot \sqrt{t} \cdot \ln t = o(t)$ with $c > 0$ an appropriately chosen constant it holds that*

$$\Pr\left(\hat{\Phi}\ \textit{unsat}\right) \leqslant \mu_{\boldsymbol{p}_n,s_c+g'(t)}(f) + o(1).$$

Under the assumption that $\Pr(\hat{\varPhi} \text{ unsat}) \geqslant \mu_{\boldsymbol{p}_n,s_c}(f) + \delta$, it follows that $\mu_{\boldsymbol{p}_n,s_c+g'(t)}(f) \geqslant \mu_{\boldsymbol{p}_n,s_c}(f) + \varepsilon$ for some constant $\varepsilon > 0$. We show that this cannot be the case under the assumption of a coarse threshold. The proof of this lemma is a simple application of Taylor's theorem and uses the fact that we evaluate the probability function at the point s_c, where $\frac{d\mu_{\boldsymbol{p},n_s}(f)}{ds} s\Big|_{s=s_c} \leqslant K$ due to Lemma 3.1.

Lemma 3.10. *It holds that* $\mu_{\boldsymbol{p}_n,s_c+g'(t)}(f) \leqslant \mu_{\boldsymbol{p}_n,s_c}(f) + o(1)$.

This contradicts our conclusion of $\mu_{\boldsymbol{p}_n,s_c+g'(t)}(f) \geqslant \mu_{\boldsymbol{p}_n,s_c}(f) + \varepsilon$ for some constant $\varepsilon > 0$. Therefore, our assumption $\Pr(\hat{\varPhi} \text{ unsat}) \geqslant \mu_{\boldsymbol{p}_n,s_c}(f) + \delta$ for some constant $\delta > 0$ has to be wrong, i.e. $\Pr(\hat{\varPhi} \text{ unsat}) \leqslant \mu_{\boldsymbol{p}_n,s_c}(f) + o(1)$. Now we can put all error probabilities together to see

$$\Pr_{x \sim \pi}\left(f(x) = -1 \mid x_T = (-1)^{|T|}\right) \leqslant \mu_{\boldsymbol{p}_n,s_c}(f) + o(1).$$

This is smaller than $\mu_{\boldsymbol{p}_n,s_c}(f) + \tau$ for sufficiently large values of n. This means, the maximally quasi-unsatisfiable subformula $\varPhi_T$ cannot be a τ-booster for any constant $\tau > 0$. Due to Lemma 3.4 the boost by every satisfiable subformula is at most as big as the one by a mqu subformula. Thus, no T which encodes a satisfiable subformula can be a τ-booster. Since we already ruled out unsatisfiable subformulas, this means there are no τ-boosters which appear with probability at least $\tau/2$. This contradicts the implication of the Sharp Threshold Theorem and therefore the assumption of a coarse threshold, thus proving Theorem 3.3. □

3.4 Relation to the clause-drawing model

After proving the sharpness of the threshold for $\mathcal{F}(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s)$ in Theorem 3.2, it now remains to relate $\mathcal{F}(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s)$ to $\mathcal{D}(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m)$.

Usually, the satisfiability threshold is only determined for the clause-drawing model and not for the clause-flipping model. Nevertheless, the following lemma shows that for certain probability distribution ensembles $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ the asymptotic thresholds of both models are the same. This allows us to determine the asymptotic threshold function of the clause-flipping model and to apply Theorem 3.2. The proofs of Lemma 3.11 and Lemma 3.12 use Lemma 2.1 and Chernoff Bounds.

Lemma 3.11. *Let* $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ *be an ensemble of variable probability distributions on* n *variables each and let* $t = \omega(1)$ *be an asymptotic threshold with respect to* m *for a monotone property* P *on* $\mathcal{D}(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m)$. *If* $\|p_n\|_2^2 = o\left(t^{-1/k}\right)$, *then* $s_c = \Theta(t)$ *is an asymptotic threshold with respect to* s *for* P *on* $\mathcal{F}(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s)$.

With the help of the former lemma, we can now prove Lemma 3.12.

Lemma 3.12. *Let* $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ *be an ensemble of variable probability distributions on* n *variables each and let* $t = \omega(1)$ *be an asymptotic threshold with respect to* s *for any monotone property* P *on* $\mathcal{F}(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s)$. *If* $\|p_n\|_2^2 = o\left(t^{-1/k}\right)$ *and if the threshold for* P *with respect to* s *on* $\mathcal{F}(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s)$ *is sharp, then* P *has a sharp threshold on* $\mathcal{D}(n, k, \boldsymbol{p}, m)$ *at* $m_c = \Theta\left(t\right)$.

Theorem 3.3, now follows from the two lemmas above and from Theorem 3.2.

Theorem 3.3. *Let $k \geqslant 2$, let $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ be an ensemble of variable probability distributions on n variables each and let $m_c = t(n) = \omega(1)$ be the asymptotic satisfiability threshold for $\mathcal{D}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m\right)$ with respect to m. If $\|\boldsymbol{p}_n\|_\infty = o\left(t^{-\frac{k}{2k-1}} \cdot \log^{-\frac{k-1}{2k-1}} t\right)$ and $\|\boldsymbol{p}_n\|_2^2 = \mathcal{O}\left(t^{-2/k}\right)$, then satisfiability has a sharp threshold on $\mathcal{D}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m\right)$ with respect to m.*

3.5 Example Application of the Theorem

We can now use Theorem 3.3 as a tool to show sharpness of the threshold for non-uniform random k-SAT with different probability distributions on the variables. As an example, we apply the theorem for an ensemble of power-law distributions.

Corollary 3.2. *Let $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ be an ensemble of general power-law distributions with the same power-law exponent $\beta \geqslant \frac{2k-1}{k-1} + 1 + \varepsilon$, where $\varepsilon > 0$ is a constant and $\boldsymbol{p}_n$ is defined over n variables. For $k \geqslant 2$ both $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s\right)$ and $\mathcal{D}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m\right)$ have a sharp threshold with respect to s and m, respectively.*

Proof. From [24] we know that the asymptotic threshold for $\mathcal{D}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m\right)$ is at $m = \Theta(n)$ for $\beta \geqslant \frac{2k-1}{k-1} + \varepsilon$. It is now an easy exercise to see that

$$\|\boldsymbol{p}_n\|_2^2 = \sum_{i=1}^{n} p_{n,i}^2 = \begin{cases} \mathcal{O}\left(n^{-2(\beta-2)/(\beta-1)}\right) & ,\beta < 3 \\ \mathcal{O}\left(\frac{\ln n}{n}\right) & ,\beta = 3 \\ \mathcal{O}\left(n^{-1}\right) & ,\beta > 3 \end{cases}$$

and that $\|\boldsymbol{p}_n\|_\infty = \max_{i=1,2,\ldots,n}(p_{n,i}) = \mathcal{O}(n^{-(\beta-2)/(\beta-1)})$. One can now verify $\|\boldsymbol{p}_n\|_2^2 = \mathcal{O}(n^{-2/k})$ and $\|\boldsymbol{p}_n\|_\infty = o(n^{-\frac{k}{2k-1}} \cdot \log^{-\frac{k-1}{2k-1}}(n))$ for $\beta > \frac{2k-1}{k-1} + 1 + \varepsilon$ and $k \geqslant 2$. Lemma 3.11 now states that the asymptotic satisfiability threshold for $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s\right)$ is at $s = \Theta(n)$. Theorem 3.2 and Theorem 3.3 now imply a sharp threshold for $\mathcal{F}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, s\right)$ and $\mathcal{D}\left(n, k, (\boldsymbol{p}_n)_{n\in\mathbb{N}}, m\right)$.

4 Discussion of the Results

In this work we have shown sufficient conditions on the variable probability distribution of non-uniform random k-SAT for the satisfiability threshold to be sharp. The main theorems can readily be used to prove sharpness for a wide range of random k-SAT models with heterogeneous distributions on the variable occurrences: If the threshold function is known asymptotically, one only has to verify the two conditions on the variable distribution.

We suspect that it is possible to generalize the result to demanding only $\|\boldsymbol{p}\|_\infty = o\left(t^{-1/k}\right)$, since the additional factor is only needed in Lemma 3.8. In any case it would be interesting to complement the result with matching conditions on coarseness of the threshold.

We hope that our results make it possible to derive a proof in the style of Ding et al. [20] for certain variable probability ensembles with a sharp threshold, effectively proving the satisfiability threshold conjecture for these ensembles.

References

1. Achlioptas, D., Kirousis, L.M., Kranakis, E., Krizanc, D.: Rigorous results for random (2+p)-sat. Theor. Comput. Sci. **265**(1-2), 109–129 (2001)
2. Alistarh, D., Sauerwald, T., Vojnović, M.: Lock-free algorithms under stochastic schedulers. In: 34th Symp. Principles of Distributed Computing (PODC). pp. 251–260 (2015)
3. Ansótegui, C., Bonet, M.L., Giráldez-Cru, J., Levy, J.: The fractal dimension of SAT formulas. In: 7th Intl. Joint Conf. Automated Reasoning (IJCAR). pp. 107–121 (2014)
4. Ansótegui, C., Bonet, M.L., Giráldez-Cru, J., Levy, J.: On the classification of industrial SAT families. In: 18th Intl. Conf. of the Catalan Association for Artificial Intelligence (CCIA). pp. 163–172 (2015)
5. Ansótegui, C., Bonet, M.L., Levy, J.: On the structure of industrial SAT instances. In: 15th Intl. Conf. Principles and Practice of Constraint Programming (CP). pp. 127–141 (2009)
6. Ansótegui, C., Bonet, M.L., Levy, J.: Towards industrial-like random SAT instances. In: 21st Intl. Joint Conf. Artificial Intelligence (IJCAI). pp. 387–392 (2009)
7. Ansótegui, C., Giráldez-Cru, J., Levy, J.: The community structure of SAT formulas. In: 15th Intl. Conf. Theory and Applications of Satisfiability Testing (SAT). pp. 410–423 (2012)
8. Bapst, V., Coja-Oghlan, A.: The condensation phase transition in the regular k-sat model. In: Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2016. pp. 22:1–22:18 (2016)
9. Boufkhad, Y., Dubois, O., Interian, Y., Selman, B.: Regular random k-sat: Properties of balanced formulas. J. Autom. Reasoning **35**(1-3), 181–200 (2005)
10. Boufkhad, Y., Dubois, O., Interian, Y., Selman, B.: Regular random k-SAT: Properties of balanced formulas. J. Automated Reasoning **35**(1-3), 181–200 (2005)
11. Bradonjic, M., Perkins, W.: On sharp thresholds in random geometric graphs. In: Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2014. pp. 500–514 (2014)
12. Bringmann, K.: Why walking the dog takes time: Frechet distance has no strongly subquadratic algorithms unless SETH fails. In: 55th Symp. Foundations of Computer Science (FOCS). pp. 661–670 (2014)
13. Chvatal, V., Reed, B.: Mick gets some (the odds are on his side). In: 33rd Symp. Foundations of Computer Science (FOCS). pp. 620–627 (1992)
14. Coja-Oghlan, A.: The asymptotic k-SAT threshold. In: 46th Symp. Theory of Computing (STOC). pp. 804–813 (2014)
15. Coja-Oghlan, A., Panagiotou, K.: The asymptotic k-SAT threshold. Advances in Mathematics **288**, 985–1068 (2016)
16. Coja-Oghlan, A., Wormald, N.: The number of satisfying assignments of random regular k-sat formulas. CoRR **abs/1611.03236** (2016)
17. Cook, S.A.: The complexity of theorem-proving procedures. In: 3rd Symp. Theory of Computing (STOC). pp. 151–158 (1971)
18. Cygan, M., Nederlof, J., Pilipczuk, M., Pilipczuk, M., van Rooij, J.M.M., Wojtaszczyk, J.O.: Solving connectivity problems parameterized by treewidth in single exponential time. In: 52nd Symp. Foundations of Computer Science (FOCS). pp. 150–159 (2011)
19. Díaz, J., Kirousis, L.M., Mitsche, D., Pérez-Giménez, X.: On the satisfiability threshold of formulas with three literals per clause. Theoretical Computer Science **410**(30-32), 2920–2934 (2009)

20. Ding, J., Sly, A., Sun, N.: Proof of the satisfiability conjecture for large k. In: 47th Symp. Theory of Computing (STOC). pp. 59–68 (2015)
21. Fortuin, C.M., Kasteleyn, P.W., Ginibre, J.: Correlation inequalities on some partially ordered sets. Communications in Mathematical Physics **22**(2), 89–103 (Jun 1971)
22. Friedgut, E.: Sharp thresholds of graph properties, and the k-SAT problem. J. Amer. Math. Soc. **12**(4), 1017–1054 (1999)
23. Friedgut, E.: Hunting for sharp thresholds. Random Struct. Algorithms **26**(1-2), 37–51 (2005)
24. Friedrich, T., Krohmer, A., Rothenberger, R., Sauerwald, T., Sutton, A.M.: Bounds on the satisfiability threshold for power law distributed random SAT. In: 25th European Symposium on Algorithms (ESA). pp. 37:1–37:15 (2017)
25. Friedrich, T., Krohmer, A., Rothenberger, R., Sutton, A.M.: Phase transitions for scale-free SAT formulas. In: 31st Conf. Artificial Intelligence (AAAI). pp. 3893–3899 (2017)
26. Giráldez-Cru, J., Levy, J.: A modularity-based random SAT instances generator. In: 24 thIntl. Joint Conf. Artificial Intelligence (IJCAI). pp. 1952–1958 (2015)
27. Giráldez-Cru, J., Levy, J.: Locality in random SAT instances. In: 26th Intl. Joint Conf. Artificial Intelligence (IJCAI). pp. 638–644 (2017)
28. Goerdt, A.: A threshold for unsatisfiability. J. Comput. Syst. Sci. **53**(3), 469–486 (1996)
29. Hajiaghayi, M.T., Sorkin, G.B.: The satisfiability threshold of random 3-SAT is at least 3.52. Tech. Rep. RC22942, IBM (October 2003)
30. Impagliazzo, R., Paturi, R.: On the complexity of k-SAT. J. Comput. Syst. Sci. **62**(2), 367–375 (2001)
31. Impagliazzo, R., Paturi, R., Zane, F.: Which problems have strongly exponential complexity? In: 39th Symp. Foundations of Computer Science (FOCS). pp. 653–663 (1998)
32. Kaporis, A.C., Kirousis, L.M., Lalas, E.G.: The probabilistic analysis of a greedy satisfiability algorithm. Random Struct. Algorithms **28**(4), 444–480 (2006)
33. Karp, R.M.: Reducibility among combinatorial problems. In: Proceedings of a symposium on the Complexity of Computer Computations, held March 20-22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York. pp. 85–103 (1972)
34. Levin, L.A.: Universal sorting problems. Problems of Information Transmission **9**, 265–266 (1973)
35. Mézard, M., Parisi, G., Zecchina, R.: Analytic and algorithmic solution of random satisfiability problems. Science **297**(5582), 812–815 (2002)
36. Mitchell, D.G., Selman, B., Levesque, H.J.: Hard and easy distributions of SAT problems. In: 10th Conf. Artificial Intelligence (AAAI). pp. 459–465 (1992)
37. Monasson, R., Zecchina, R.: Statistical mechanics of the random k-satisfiability model. Phys. Rev. E **56**, 1357–1370 (Aug 1997)
38. Monasson, R., Zecchina, R., Kirkpatric, S., Selman, B., Troyansky, L.: Phase transition and search cost in the 2+ p-sat problem. 4th Workshop on Physics and Computation, Boston, MA, 1996. (1996)
39. Monasson, R., Zecchina, R., Kirkpatrick, S., Selman, B., Troyansky, L.: 2+p-sat: Relation of typical-case complexity to the nature of the phase transition. Random Struct. Algorithms **15**(3-4), 414–435 (1999)
40. Mull, N., Fremont, D.J., Seshia, S.A.: On the hardness of SAT with community structure. In: 19th Intl. Conf. Theory and Applications of Satisfiability Testing (SAT). pp. 141–159 (2016)

41. Müller, T.: The critical probability for confetti percolation equals 1/2. Random Struct. Algorithms **50**(4), 679–697 (2017)
42. O'Donnell, R.: Analysis of Boolean Functions. Cambridge University Press (2014)
43. Rathi, V., Aurell, E., Rasmussen, L.K., Skoglund, M.: Bounds on threshold of regular random k-sat. In: 13th Intl. Conf. Theory and Applications of Satisfiability Testing (SAT). pp. 264–277 (2010)